























## **Slide 29**

So now we can add TORCH to the summary table. TORCH was a full factorial trial designed with mortality as a primary endpoint, with a three-year duration. If we contrast the design of TORCH with IMPACT, again a few things stand out. First, TORCH recruited a different COPD patient population than IMPACT. While TORCH did recruit subjects with moderate-to-very-severe COPD by spirometry, TORCH's enrollment criteria did not specifically require a history of prior COPD exacerbations. There was also no requirement for prior COPD maintenance medications. TORCH imposed a run-in period that involved discontinuation of ICS, LABA, and LAMA prior to randomization. As we go through the data for each trial, I encourage you to keep these important design characteristics in mind.

## **Slide 30**

After reviewing the design elements of the trials, we can move into describing the patients in each, starting with IMPACT. The tables you will see here are also in Section 4.2 of your briefing document.

## **Slide 31**

This table describes selected clinical features of IMPACT's randomized patient population by trial arm. As you can see, the patients in IMPACT had higher markers of severity than many previous COPD trials; they had a mean FEV1 of around 46% predicted, high St George's Respiratory Questionnaire total scores, and – importantly – a high proportion of frequent exacerbators in the prior year. And it's important to note here, that these were frequent exacerbators despite the pre-study maintenance medications that they were required to be on for study entry.

As we think about IMPACT's design, interventions, and the generalizability of its all-cause mortality results, I ask you to consider the severity of these patients, and how they would be managed in clinical practice.

## **Slide 32**

IMPACT allowed subjects on any pre-study medication to enroll. If we take a closer look at these pre-study medication regimens, we can see that 38% of the trial population was already on triple therapy with an ICS, a LABA, and a LAMA prior to enrolling in IMPACT, but still met exacerbation requirements. If we look at subjects that had any pre-study ICS-containing regimen, they made up 71% of the trial population. In IMPACT, these pre-study ICS subjects were almost all on an ICS/LABA combination or on triple therapy. In contrast, only 29% of IMPACT was ICS-naïve going into the trial. These two subgroups of pre-study ICS and ICS-naïve will be important when we talk about subgroup analyses of IMPACT.

## **Slide 33**

So, if we summarize the population of IMPACT, pursuant to the design choices and entry criteria, the IMPACT trial was successful in recruiting subjects that had uncontrolled COPD, as shown here. 70% were frequent exacerbators in the prior year. And that high proportion of prior exacerbations is present despite 71% of the subjects already receiving pre-study ICS medications, and 38% already receiving pre-study triple therapy.

### **Slide 34**

While our prior tables looked at the disease characteristics in IMPACT alone, this table – which is Table 5 in your briefing document – compares some of those metrics across IMPACT, SUMMIT, and TORCH. As you can see from the table, SUMMIT subjects had a mean FEV1 that approached 60 percent predicted, while TORCH's was 44%. A similar pattern is echoed in the SGRQ scores. SGRQ scores for subjects in SUMMIT were lowest, but those of TORCH were relatively close to IMPACT. The proportion of frequent exacerbators, however, tells a slightly different story. Here, IMPACT's 70% stands out compared to the other two trials, as does its higher percentage of subjects with severe exacerbations in the prior 12 months.

### **Slide 35**

If we look at the pre-study medications across trials using Table 6 from the briefing document, we can see that 71% of IMPACT's subjects used pre-study ICS medication, followed by 49% in the TORCH trial, and 33% in the SUMMIT trial. While not a 1-to-1 comparison, the ordering of these proportions of pre-study ICS users roughly echo the proportions of frequent exacerbators across trials that we saw on the prior table. It's worth noting that the TORCH trial was performed at a time when triple therapy was not widely available for patients so the comparison across triple therapy subgroups does not necessarily follow severity patterns.

### **Slide 36**

Since we mentioned pre-study medications, it's worth introducing the idea that pre-study therapy has been a point of controversy in the literature surrounding COPD trials for years.

Suissa and colleagues made this point, among others, in an article published in 2008. While we will discuss this more later in the presentation, the controversy surrounds pre-study medications, and the interpretation of data from subjects who enter a randomized trial looking to answer a question about a specific drug, despite some of them already being on that drug or a similar drugs in the class. The authors noted that such a design choice led to separate scenarios at randomization, using ICS as an example. For those subjects without pre-study ICS – meaning those who were ICS-naïve prior to the trial – randomization to an ICS-containing trial arm represented the addition of an ICS, while randomization to an arm that did not contain inhaled corticosteroids did not, and acted as a control; this design is reminiscent of a standard placebo-controlled trial design for an add-on drug.

On the other hand, for those entering the trial on pre-study ICS, randomization represents something different. For these patients, randomization to an ICS-containing trial arm does not add an ICS, but could be considered continuation of ICS therapy. In the same vein, randomization to an arm that does not contain ICS could be considered ICS removal, since the subject used this drug class prior to study entry. A more detailed discussion of these concepts is discussed in Section 4.5.1 of your briefing document.

In this same article, using exacerbation data from prior trials, the authors provided analyses that suggested that the proportion of subjects with pre-study ICS entering a trial was associated with the relative risk difference observed in that trial. Suissa and colleagues considered the issue of pre-study medication use a methodologic flaw and stated that results from trials that mixed these pre-study medication groups and their corresponding different interventions might not be interpretable, due to

the issues described. This concept is pertinent to our discussion of the subgroups in the IMPACT trial and I will expand on this discussion more after the statistical review by Ms. Duke.

### **Slide 37**

But, if we turn back to the population data from the trials at hand, we can contrast IMPACT with SUMMIT and TORCH in our table. We see that IMPACT had a higher proportion of frequent exacerbators and a higher proportion of pre-study ICS users than either of the other trials.

### **Slide 38**

In summary, despite a need for therapies that improves mortality in COPD, there are currently no COPD therapies approved with an indication or labeling claim for improving all-cause mortality. Previous trials failed to show a difference for their ICS/LABA versus placebo comparisons.

The role of ICS in COPD is still somewhat controversial, but involves balancing the recognized exacerbation benefit with the risk of pneumonia in different COPD patient populations. To date, trials involving inhaled corticosteroids have not demonstrated a mortality benefit attributable to the ICS component.

The TRELEGY ELLIPTA COPD development program focused on providing evidence of a difference in exacerbations for the triple combination versus its dual components while enrolling a population with any pre-study medication use that might allow for expansion of the indication. Neither the Sponsor nor the FDA raised concerns about ICS removal during trial design and development.

### **Slide 39**

If we summarize IMPACT's design and contrast it with SUMMIT and TORCH, IMPACT was designed as an exacerbation trial that designated all-cause mortality as an "other" endpoint, and the submitted all-cause mortality analyses included post hoc data collection. In contrast, TORCH and SUMMIT were designed with all-cause mortality as a primary objective. IMPACT evaluated all-cause mortality over one year, while SUMMIT and TORCH evaluated over longer periods.

We can also note that IMPACT's enrollment criteria recruited a sicker population of COPD patients, and these patients had uncontrolled COPD despite maintenance medications. TORCH and SUMMIT's inclusion criteria did not require the same markers of disease severity. IMPACT required pre-study medication use for enrollment, and IMPACT's run-in continued pre-study medications until the day of randomization. TORCH and SUMMIT did not have a requirement for pre-study medication use, and both required changes to any existing pre-study medication regimens prior to randomization through run-ins and other design elements.

Finally, if we look at the patient populations in each trial 71% of IMPACT's randomized population had pre-study medication regimens that included ICS, and 38% were on pre-study triple therapy. 70% of subjects in IMPACT were frequent exacerbators at trial entry despite the use of these pre-study medications. In contrast, only about a third of SUMMIT's randomized population used pre-study ICS, and about half of TORCH's randomized population used pre-study ICS.

We ask you to keep these points in mind as you consider the efficacy results.

**Slide 40**

This concludes the Overview of the Clinical Program. With this portion complete, I will pass the presentation along to my statistical colleague, Ms. Susan Duke, who will present the Statistical Review of Efficacy.

## **August 31, 2020 Pulmonary-Allergy Drugs Advisory Committee Meeting**

### **Script for FDA Presentation 2: Statistical Review of Efficacy**

**FDA Presenter: Susan Duke, MS, MS Statistical Reviewer**

#### **Slide 1**

My name is Susan Duke. In this presentation, I am going to describe important aspects of the **statistical** review of this supplemental NDA for an All Cause Mortality labeling claim for Trelegy Ellipta.

#### **Slide 2**

Here is the outline of this talk: I'll discuss the IMPACT study and all-cause mortality, independent supportive evidence, exploratory analyses and then summarize.

#### **Slide 3**

In this first section I will discuss the IMPACT study design features, analysis plan, follow-up for mortality and overall results.

#### **Slide 4**

The IMPACT study was designed to understand the contribution of FF and UMEC to FF/UMEC/VI with respect to exacerbations. It was not designed to assess all cause mortality as a primary or secondary objective. The primary endpoint was annual rate of moderate to severe exacerbations. Secondary endpoints were FEV<sub>1</sub>, SGRQ, and time to first exacerbation. All-cause mortality was among the many exploratory or 'other' endpoints.

Because the study was designed for exacerbations, it was not powered for mortality. Study duration was 1 year, which was reasonable for exacerbations. Mortality studies for COPD have utilized a longer duration.

#### **Slide 5**

Regarding the IMPACT analysis plan, there was a plan to control Type I error across the pre-specified analyses of primary and secondary endpoints. As already noted, IMPACT met its primary endpoint, demonstrating a statistically significant effect on the rate of moderate to severe exacerbations from the comparisons of FF/UMEC/VI versus both double combinations. Furthermore, all secondary analyses were statistically significant.

However, as mentioned, All-cause Mortality was not a primary or secondary endpoint but rather was one of many 'other' endpoints, most of which had two pairwise comparisons, and none of which were under Type 1 error control. This makes interpretation of the results challenging. With every additional exploratory analysis conducted, the probability of observing p-values less than 0.05 purely due to chance increases. In addition, such exploratory analyses may be subject to substantial random high bias.

The analysis of all-cause mortality was a Cox proportional hazards model, with covariates of gender and age, evaluating the effect of UMEC by comparing FF/UMEC/VI to FF/VI, and evaluating the effect of FF by comparing FF/UMEC/VI to UMEC/VI.

#### **Slide 6**

The analyses conducted for this submission included different degrees of mortality follow-up.

The on-treatment analysis includes deaths and follow-up while subjects were on study medication and excludes deaths and follow-up that occurred after treatment discontinuation. The on-study analysis includes deaths and follow-up that occurred while subjects remained on the study, regardless of whether they continued study treatment, but excludes any mortality events and follow-up that occurred after subjects withdrew from the study. The All vital status follow-up analysis includes all available deaths and mortality follow-up over the 52-week study, including additional vital status follow-up in patients who withdrew from the study that was obtained by the applicant after the study completed to inform this application.

The FDA focus is on the analyses including all vital status follow-up, also termed (“ITT + VS + VSFU”) in your briefing document. We are primarily interested in evaluating the difference in survival regardless of adherence and use of other therapy. Therefore, we focus on the analysis that provides the most complete data, with nearly all vital status follow-up included, in order to provide what we believe are the most reliable results. Analyses that include different degrees of follow-up produce slightly different quantitative results from the analyses including all vital status follow-up that I present here, but the key conclusions and uncertainties are similar regardless of the analysis that is emphasized.

#### **Slide 7**

This table displays the percentages of patients with complete all-cause mortality through 52 weeks in the analyses that included only on-treatment follow-up, only on-study follow-up, or included all available vital status follow-up. The analysis including all vital status follow-up includes considerably more complete information, with 99.6% of patients having complete vital status follow-up through Week 52. From this point forward, our presentations will focus on the all vital status follow-up analysis unless otherwise noted.

#### **Slide 8**

This table shows the results from the all-cause mortality analysis. The comparison between FF/UMEC/VI versus FF/VI, which isolates the effect of UMEC, did not show an effect, with an estimated hazard ratio of 0.89. The comparison of FF/UMEC/VI versus UMEC/VI, which isolates the effect of FF, and is the primary basis of the proposed claim by the applicant, revealed an estimated hazard ratio of 0.72, with a 95% CI of 0.53 to 0.99 and a p-value of 0.042. While this p-value is less than the commonly used threshold of 0.05, this was an exploratory analysis that was not included in the multiple testing strategy to control Type I error, so results are difficult to interpret. Furthermore, and as will be discussed in greater detail later, there is often an expectation of considerably more persuasive evidence of an effect when a proposed effectiveness claim relies on results from a single study.



## **Slide 9**

This Kaplan Meier plot uses the same data and shows the probability of death over the 52-week study, with the y-axis ranging from 0 to 3.5%. The blue line represents the FF/UMEC/VI treatment arm, the red line represents FF/VI and the uppermost green line, which has separated from the other two treatment arms, is the UMEC/VI treatment arm.

The early separation of the mortality curves was unexpected and may not be consistent with previous trials that evaluated inhaled corticosteroids in COPD based on the literature, including two trials that I will discuss next. I will discuss this timeframe of efficacy and the uncertainties associated with it in more detail later.

## **Slide 10**

So that concludes my discussion of the design and the overall results from the all-cause mortality analysis of the IMPACT study. Given that the Applicant's proposed claim is based on results from a single trial that was not designed to evaluate mortality, we thought it was important to consider whether any relevant independent data provide corroborating evidence to support the purported effect of fluticasone on survival. In that light, next we turn our attention to independent supportive evidence from the SUMMIT and TORCH studies.

## **Slide 11**

This table compares characteristics of the three studies that Dr. Busch presented earlier. We acknowledge that SUMMIT and TORCH involved different populations and different comparisons to evaluate the ICS effect than IMPACT, and that the fluticasone study drug evaluated in TORCH was fluticasone propionate. Nevertheless, these data provide the most reliable independent data to help inform the proposed claim based on IMPACT and are considered critical in light of the uncertainties about the persuasiveness of the results from that single study.

## **Slide 12**

SUMMIT and TORCH were both designed with a primary objective of evaluating mortality and had primary analyses that were powered to detect differences in mortality events between ICS/LABA and placebo. And each was for a longer duration than IMPACT: The SUMMIT trial was an event-driven trial, with a median follow-up of 1.8 years and a maximum duration of 46 months. TORCH was 3 years in duration.

As mentioned previously, neither of the two studies achieved statistical significance in the primary analysis of ICS/LABA vs. placebo with respect to all-cause mortality. Given that our goal in looking at information from SUMMIT and TORCH was to seek supportive evidence for an effect of fluticasone on mortality, we primarily focus on the comparisons that isolate the ICS effect. You will see in the next slide that SUMMIT and TORCH did not demonstrate statistically significant differences in all-cause mortality for either the comparison of ICS/LABA vs LABA or ICS vs placebo.

### **Slide 13**

Here are the treatment comparisons from SUMMIT and TORCH that isolate the effect of fluticasone.

As shown here, SUMMIT and TORCH had between 398 and 526 events for each of the relevant fluticasone comparisons. In contrast, IMPACT's FF/UMEC/VI vs UMEC/VI comparison yielded only 164 mortality events, so each comparison in SUMMIT and TORCH provided roughly three times the amount of statistical information as IMPACT and therefore had greater power than IMPACT to detect a difference if it existed.

As illustrated by the hazard ratios and confidence intervals for the ICS/LABA vs LABA and ICS vs placebo comparisons, SUMMIT and TORCH did not provide evidence of an effect of the fluticasone product on mortality. Therefore, neither SUMMIT or TORCH offered corroborating evidence for the results found in IMPACT.

### **Slide 14**

Here are the survival curves for SUMMIT and TORCH over 3 years. The x-axis ranges from 0 to 3 years, with the y-axis showing the probability of mortality and ranging from 0 to 20%. One can see similarities between the two mortality studies, with a lack of clear separation between the placebo, single component, and ICS/LABA combination arms over time.

### **Slide 15**

Now let's compare only the first year of the SUMMIT and TORCH results with the 1 year IMPACT study. The y-axis has changed in this visualization, and now shows the probability of mortality and ranges from 0 to 4.5%, and the x-axis spans 0 to 365 days. The separation of curves was strikingly different in IMPACT, with early separation of the UMEC/VI arm, designated here by the green line.

This difference in behavior of the UMEC/VI arm in IMPACT warranted further investigation.

### **Slide 16**

In that light, I next discuss 2 exploratory analyses of IMPACT, analyses to further explore the timeframe of the observed mortality separation and subgroup analyses by pre-study ICS status. It is important to note that any exploratory analysis is, at most, hypothesis generating.

### **Slide 17**

Exploratory analyses using time-to-event visualizations and proportional hazards models stratified by time suggested that the separation in mortality between UMEC/VI and FF/UMEC/VI in IMPACT occurred within the first 90 days. For example, here we show the Kaplan-Meier plot on the left for the full 52 weeks of IMPACT, and you can see the early division between curves. But how did the curves behave if that first 90 days of mortality data were removed? If we remove the first 90 days, and look only at mortality occurring after Day 90, there was no difference among the curves. These analyses are exploratory, and analyses after Day 90 condition on the post-randomization variable of surviving to Day 90 and may be subject to bias. Nevertheless, these additional results suggest that the observed difference at trial completion was driven by early events. If interpreted as a benefit, this early signal for mortality is unexpected, given multi-year data from previous trials such as TORCH and SUMMIT. Furthermore, if an effect of FF on mortality relied on a mechanism such as prevention of severe

exacerbation events the timeframe of efficacy might be expected to follow a pattern of gradual accumulation, since severe exacerbations are rare events.

There are many potential explanations for the observed early separation of curves, including a real benefit or a chance finding. Another potential explanation revolves around the potential effect of ICS removal, as Dr. Busch will discuss later. I will next discuss the results and statistical interpretation of our exploratory analysis of pre-study ICS subgroups.

### **Slide 18**

Here we show stratified Kaplan-Meier plots of IMPACT by pre-study ICS status. The figure on the left shows mortality over time in the subset of patients who were on pre-study ICS therapy. The figure on the right shows mortality over time in the subset of patients who were ICS-naïve prior to the study. The differences in the two Kaplan Meier curves are striking. The results on the left in subjects with pre-study ICS shows a higher early and total probability of death events in the UMEC/VI arm. In contrast, the results on the right in subjects who were not on a pre-study ICS does not show any separation between the curves. The ICS naïve subgroup analysis provides information relevant to clinical practice about the possible benefit of adding FF. While this subgroup comprises only 29% of the trial population, and was underpowered to detect a difference in mortality, the lack of a trend towards a mortality difference in the FF/UMEC/VI versus UMEC/VI comparison is notable.

We also tested for the presence of a statistical interaction between pre-study ICS and treatment by adding pre-study ICS and treatment by pre-study ICS interaction terms to the all-cause mortality analysis model. The pre-study ICS interaction term for the FF/UMEC/VI vs UMEC/VI pairwise comparison had a p-value of 0.08. These results, while exploratory in nature, suggest that the overall population mortality results may be difficult to interpret, and that it may be more appropriate to analyze the pre-study ICS subgroups separately.

Among other explanations, the results from this pre-study ICS subgroup analysis could be attributed to randomization-mandated ICS removal that occurred in the UMEC/VI arm, a topic which Dr. Busch will explain and discuss in further detail in his next presentation.

### **Slide 19**

I will now summarize our statistical findings, beginning with a reminder of FDA effectiveness standards.

### **Slide 20**

FDA standards for claims of effectiveness are described in the guidances for determining substantial evidence noted at the bottom of this slide.

The gold standard is evidence from at least 2 adequate and well-controlled studies.

Otherwise, in some specific settings, a finding of substantial evidence of effectiveness to support a claim can be made based on “one adequate and well-controlled clinical investigation plus confirmatory evidence”. Key factors to allow for such a determination include the “persuasiveness of evidence from a single study” and the “robustness of confirmatory evidence.” The guidance indicates that reliance on a single study should “be limited to situations in which the trial has demonstrated a clinically meaningful and statistically very persuasive effect ...” In that light, we note there is often an expectation of

evidence of an effect at a statistical significance level considerably lower than 0.05 when a proposed effectiveness claim relies on results from a single study.

I will come back to this guidance as I wrap up on my next and final slide.

## **Slide 21**

When we look at the totality of statistical evidence, there are a number of uncertainties that warrant the consideration and discussion of the Advisory Committee. The first question is about the persuasiveness of the IMPACT all-cause mortality results. IMPACT was not designed for mortality and had a relatively short duration of 1 year, with no Type I error control for the mortality analysis. We acknowledge that all analyses of primary and secondary endpoints were statistically significant, that mortality is a clinically important outcome, and that many (but not all) of the exploratory analyses of “other” endpoints had nominal p-values below the 0.05 threshold. Nevertheless, the mortality evaluation was one of a long list of exploratory analyses. It would not be unusual to find nominal p-values below 0.05 just by chance when evaluating multiple exploratory endpoints, and such analyses may also be subject to substantial random high bias. Furthermore, there are questions about the strength of evidence with a p-value of 0.042 from only a single study.

The second question is about the degree of supportive evidence from independent data, given the effectiveness standards I described on the previous slide and given that the proposed claim primarily relies on a single study. The longer SUMMIT and TORCH trials were specifically designed for All Cause Mortality, had a longer duration, and included roughly 3-fold more events than IMPACT. However, these studies did not provide evidence of mortality effects for the ICS/LABA product in their primary analyses, nor did they provide corroborating evidence of a mortality benefit for fluticasone products in COPD based on comparisons that isolated the effect of FP in TORCH and FF in SUMMIT.

And finally, exploratory analyses of IMPACT add additional uncertainty, in terms of the unexpected early separation of survival that was observed and the subgroup analysis results suggesting that the observed survival difference was driven by the subset of patients who were taking ICS pre-study. Dr. Busch is going to discuss potential implications of these exploratory results and the additional uncertainties they convey in greater detail in the next presentation.

## **August 31, 2020 Pulmonary-Allergy Drugs Advisory Committee Meeting**

### **Script for FDA Presentation 3: Clinical Considerations**

**FDA Presenter: Robert Busch, MD, Medical Officer**

#### **Slide 1**

Thank you, Ms, Duke. So now that we've discussed the efficacy results, we can move to a discussion of the clinical considerations for these data and the application as a whole.

#### **Slide 2**

In doing so, I will first talk about pre-study medication considerations including extant data on ICS removal and then present exploratory subgroup analyses from IMPACT, SUMMIT, and TORCH defined by pre-study medication use, with a focus on the pre-study ICS and ICS-naïve subgroups.

After that, I'll start bringing the concepts that we've presented together by discussing the uncertainties in the interpretation of the all-cause mortality results, including the statistical persuasiveness of the data from IMPACT, the evidence for an ACM benefit for fluticasone across trials, the timeframe of efficacy observed in IMPACT, interpretations of the pre-study ICS subgroup data under an ICS removal paradigm, and the generalizability of IMPACT's data and results to clinical practice.

And, finally, I'll provide a summary of the efficacy results in context.

#### **Slide 3**

Before jumping in to the subgroups, I'm going to first set the stage a bit by talking about the general topic of ICS removal in COPD.

#### **Slide 4**

There have been concerns and controversy around the degree of negative effects due to ICS removal in COPD for at least 2 decades, if not more.

Early data on ICS removal included non-randomized designs that observed the effects of ICS removal during the run-in to a randomized trial. An example of this is the study by Jarad and colleagues from 1999, that observed the effects of ICS removal during enrollment and run-in to the ISOLDE trial. 272 patients who had been clinically stable for 6 weeks were enrolled and going through an 8-week run-in that allowed only short-acting bronchodilator use; this meant that the 160 subjects with pre-study ICS were required to remove ICS for this run-in period, while the remaining 112 ICS-naïve subjects did not. The authors reported that 38% of the pre-study ICS subjects had an exacerbation during the run-in. In contrast, only 11% of the ICS-naïve subjects had an exacerbation. While these data were observed using a non-randomized design, they raised concerns about the effects of ICS removal, and they give us a framework for examining the run-in data on mortality and ICS removal from the TORCH trial, as presented in your briefing document.

Another trial, the COSMIC trial from GSK published by Wouters and colleagues in 2005, recruited 497 subjects with COPD and standardized their ICS/LABA treatment over 3 months. Then these same

subjects were randomized to just LABA (meaning ICS removal) versus continuation of ICS/LABA. Wouters and colleagues summarized the observed results of the trial in their paper entitled “Withdrawal of fluticasone propionate from combined salmeterol/fluticasone treatment in patients with COPD causes immediate and sustained disease deterioration: a randomized controlled trial”. The data suggested that ICS removal led to a decrease in lung function and symptom scores after ICS removal, along with a deterioration in some measures of COPD exacerbations. It is notable that the conclusion in this article supported a sustained disease deterioration after ICS removal.

Van der Valk and colleagues as well as Choudhury and colleagues, among others, also conducted studies that further supported negative effects of ICS removal in COPD, including effects on exacerbations. None of these studies, however, were designed or powered to assess for a mortality difference.

### **Slide 5**

There are many other studies that addressed ICS removal to a varying degree, some of which are summarized in your briefing document under Section 4.5.1, but I am going to briefly discuss three trials designed around randomized ICS removal here. These trials are presented for context, but also because they – and studies like them – provide us with a paradigm for interpreting data after randomized removal of ICS.

The first of those trials is the WISDOM trial, published by Magnussen and colleagues in 2014. This trial recruited COPD patients with 1 or more exacerbations of any severity in the prior year, standardized their treatment to ICS/LABA/LAMA for 6 weeks, and then randomized them to a tapered ICS removal or continuation of triple therapy. Importantly, the majority – if not all – of these subjects were not frequent exacerbators.

Graphically, we can map this trial using a schematic similar to IMPACT, SUMMIT, and TORCH. Any maintenance COPD regimen was allowed pre-study, a run-in was imposed on ICS/LABA/LAMA and then ICS was removed in a randomized fashion using a 6-week taper. Among this patient population, subjects in the tapered ICS removal arm experienced a rate of moderate-to-severe exacerbations that was within the noninferiority margin compared to those who continued ICS. However, despite recruiting a patient population in which over 60% of patients only reported one exacerbation of any severity in the prior year, a trend towards increased severe exacerbations was observed in the ICS removal arm of this trial, as conveyed in a letter to the editor. In their reply, Magnussen and colleagues acknowledged the observed numerical increase in severe exacerbations, and opined that this was a transient increase after ICS were completely removed. And while this trial was not powered to assess mortality, a numerically higher rate of fatal adverse events was observed in the ICS removal group compared to the ICS-continuation group.

Importantly, the WISDOM trial provides an example of the interpretation of data from a study design in which the study intervention is a randomized removal of ICS. The effects – or lack of effects - observed in the WISDOM trial is attributed to an intervention of an ICS removal event, with ICS-continuation serving as the control. This paradigm is important to consider, and while IMPACT’s enrolled population was very different based on baseline characteristics, we will consider IMPACT subgroup data presenting the potential effect of ICS removal in a similar way.

## **Slide 6**

Two additional trials that examined the effect of ICS removal in relatively controlled COPD patients are INSTEAD and SUNSET, by Rossi and colleagues and Chapman and colleagues, respectively, and I've set up schematics for each of them here as additional examples. Both SUNSET and INSTEAD enrolled subjects on an ICS-containing therapy (ICS/LABA for INSTEAD, triple therapy for SUNSET), standardized therapy during a run-in period, and then randomized to ICS removal or ICS continuation. These trials recruited COPD patients who were not frequent exacerbators, and INSTEAD specifically recruited subjects with a history of no exacerbations in the prior year, both cases where ICS removal might be considered in practice. As with WISDOM, I am reviewing the design of these ICS removal trials because they were designed to evaluate for a potential negative effect of ICS removal among specific patient populations where the added benefit of ICS use was unclear, and the results were interpreted as an ICS removal intervention compared to a control of ICS continuation. As we move into subgroup analyses of IMPACT and other studies, please keep this in mind.

## **Slide 7**

So, while we have examples of studies and randomized trials that show a symptomatic decline in COPD after ICS removal, there is still controversy regarding the degree of effects of ICS removal. Lung function and PRO data consistently suggest a decline after ICS removal, but the data on exacerbations endpoints are less straightforward, and may depend on the patient population examined. Data among COPD patient groups with a history of exacerbations may suggest a trend towards higher rates of exacerbations after ICS removal (including severe exacerbations), while data among more controlled COPD patient groups may suggest non-inferiority on exacerbation endpoints. However, none of these trials examined a patient population with the degree of disease severity that was seen in IMPACT.

Importantly, WISDOM, SUNSET, and INSTEAD were trials that used ICS removal events to answer safety questions about ICS removal, and to suggest that the negative effects might be acceptable in the risk-benefit evaluation of certain patients with controlled COPD. It is not clear that the study interventions from these trials could provide evidence to support a claim about ICS addition. The study intervention in these three trials took patients with pre-study ICS-containing therapies and randomized them to ICS removal versus ICS-continuation as a control; no subject had ICS added as part of randomization. This is fundamentally a different paradigm than looking at the effect of adding an ICS versus placebo among ICS-naïve patients.

## **Slide 8**

Now that we've discussed ICS removal in the literature, we can talk about how the paradigm we saw in WISDOM and other studies may apply to studies that enrolled patients on any pre-study medication

## **Slide 9**

To do that, I refer back to the article by Suissa and colleagues that used ICS as an example, which I mentioned in my first presentation. In many COPD trials, including IMPACT, SUMMIT, and TORCH, subjects in the trial could be ICS-naïve, or could have pre-study ICS, and may be randomized to therapies with or without ICS components. This means that, fundamentally, there are different functional interventions that occur at randomization, which I will demonstrate here.

As you can see in this schematic, ICS-naïve subjects entering the trial can be randomized to either the addition of ICS (which serves as the intervention) or placebo (which serves as the control). This design is familiar from traditional add-on trial designs.

In contrast, the situation for subjects on pre-study ICS is fundamentally different. Subjects already on pre-study ICS can't have inhaled corticosteroids "added" at randomization, although we acknowledge that the ICS product may change. As you can see here, pre-study ICS subjects can have the ICS removed by randomization, or they can continue ICS. Since their ICS status has not changed, the arm that continues ICS could be considered as a control arm. Since ICS continuation acts as the control, the subjects that were randomized to ICS removal act as the intervention, similar to the paradigm in WISDOM and the other ICS removal trials we looked at earlier.

These two design paradigms – ICS addition for ICS-naïve subjects, and ICS removal for pre-study ICS subjects – are fundamentally different for the subjects going through a clinical trial; however many COPD trials – including IMPACT, SUMMIT, and TORCH – have presented results that simply combine the results of these subgroups into an overall effect estimate. Suissa and colleagues suggested that effect estimates combining data from these two subgroups were uninterpretable. Suissa, however, primarily analyzed data for COPD exacerbations, not mortality.

#### **Slide 10**

Now that we've discussed the conceptual issues with pre-study therapy, we can take another look at IMPACT using subgroups based on pre-study therapies.

#### **Slide 11**

If we apply these principles to IMPACT – first looking at the pre-study triple therapy subgroup and the same type of schematic that I've shown previously as an example – we can see that IMPACT imposed ICS removal events on some patients at randomization. Pre-study triple therapy subjects entered the IMPACT trial already using ICS, LABA, and LAMA, so these patients had no chance to have any therapeutic modality added by randomization. These pre-study triple therapy subjects continued all three therapies all the way up to randomization. But upon randomization, subjects randomized to UMEC/VI (the LAMA/LABA arm in the schematic) underwent an intervention of ICS removal. In contrast, subjects randomized to TRELEGY ELLIPTA (the ICS/LABA/LAMA arm) did not have any therapeutic modality added, so they could reasonably be considered the control arm in this case.

We can see that this pre-study triple therapy schematic shows similarities to the design of the WISDOM and SUNSET trials, and we end up looking at the effect of ICS removal in terms of safety. However, there are a few important differences between these trials: subjects in the IMPACT trial had uncontrolled and more symptomatic COPD compared to those other ICS removal trials, and the pre-study triple therapy group in IMPACT comprised over 3900 subjects, of which about 20% were randomized to UMEC/VI through ICS removal.

#### **Slide 12**

The all-cause mortality results for this pre-study triple therapy subgroup are pictured here in the left panel as a Kaplan-Meier curve, while the right panel represents the all-cause mortality results in subjects without pre-study triple therapy. In each case, the TRELEGY ELLIPTA arm is the blue curve, and the



UMEC/VI arm is the green curve. The FF/VI curve, in red, is included for completeness. The x-axis is time, and the y-axis is the probability of all-cause mortality. Please note that the scale of the y-axis in this graph only spans 0 to 3.5%, rather than 0 to 100%.

If we first focus on the pre-study triple therapy subgroup, shown in the left panel, we can see that the UMEC/VI arm and the TRELEGY ELLIPTA arm suggest a difference in mortality that develops in approximately the first 90 days. But, because each of these subjects already had pre-study triple therapy, the green UMEC/VI arm could be considered the interventional arm, with the intervention being ICS removal. Under this same paradigm, the blue TRELEGY ELLIPTA arm could be considered the control arm, since subjects in this arm continued on ICS/LABA/LAMA and randomization didn't add a drug class. Because the blue arm is considered a control arm, it is unclear whether we should attribute a benefit to TRELEGY ELLIPTA here. Instead, under this paradigm, these data suggest that the intervention of ICS removal led to increased mortality events among these COPD subjects.

And if we look at the right panel, we have to consider that subjects without pre-study triple therapy might still have had a pre-study ICS (for example, subjects taking pre-study ICS/LABA, without a LAMA). So while the pre-study triple therapy group is helpful as an example, if we want to understand the full extent of ICS removal in the data, we would need to look at the subgroup of subjects who had any pre-study ICS therapy.

### **Slide 13**

To do that, we can go back to our familiar schematic and define a subgroup of subjects whose pre-study therapy contained ICS in any form. In IMPACT, this pre-study ICS subgroup was comprised mostly of subjects that used pre-study ICS/LABA or those with pre-study triple therapy.

If we look at this pre-study ICS subgroup, we can see that subjects with pre-study ICS entered the trial and continued their ICS through the run-in period. But upon randomization, subjects randomized to UMEC/VI (the LABA/LAMA arm in the schematic) underwent an intervention of ICS removal. In contrast, subjects randomized to TRELEGY ELLIPTA (the ICS/LABA/LAMA arm) or to the FF/VI (the ICS/LABA arm) continued ICS. However, only if we look at the LABA/LAMA to ICS/LABA/LAMA comparison, marked here, can we get a valid factorial comparison of the effect of the ICS component, since this is the only comparison where all other factors – including the potential addition of a LAMA, for example – are equal between the groups. This comparison isolates the contribution of the ICS, and in the case of this subgroup, of the removal of ICS.

As we saw with triple therapy, this ICS removal schematic is similar to other ICS removal trials we have discussed, but IMPACT recruited a very different patient population. And in terms of size, IMPACT's pre-study ICS subgroup comprises 71% of the subjects in IMPACT, meaning this subgroup contains over 7000 subjects, and over 1400 were randomized to ICS removal in the UMEC/VI arm.

### **Slide 14**

If we look at these results in the same way we presented the pre-study triple therapy results, as in Ms. Duke's previous presentation, we get this. The all-cause mortality results for this pre-study ICS subgroup are pictured here in the left panel as a Kaplan-Meier curve, while the right panel represents the all-cause mortality results in subjects who were ICS-naïve. Again, the TRELEGY ELLIPTA curve is in blue, and

UMEC/VI is the green curve. The FF/VI curve, in red, is included for completeness. The x-axis and y-axis are unchanged from the previous graphs, and again, the y-axis in this graph only spans 0 to 3.5%.

If we look at the all-cause mortality results for this pre-study ICS subgroup in the left panel, we can see that the UMEC/VI arm and the TRELEGY ELLIPTA arm suggest a difference in mortality that develops in approximately the first 90 days. But, again, since each of these subjects already had pre-study ICS, subjects in the green UMEC/VI arm had an ICS removal intervention at randomization, and this could be considered the interventional arm. And again, the blue TRELEGY ELLIPTA arm could be considered the control arm, since subjects in this arm continued ICS. Similar to the results we saw with pre-study triple therapy, under this paradigm, these data suggest that ICS removal may have led to increased mortality events among these COPD subjects who were already uncontrolled at the time of enrollment despite taking an ICS medication.

And, when we look at tables of these data in a moment, we have to consider how best to present the hazard ratios under a paradigm where ICS-removal is the intervention and ICS continuation is the control. Instead of presenting hazard ratios as TRELEGY ELLIPTA versus UMEC/VI, with UMEC/VI considered the control, it may be reasonable to present “flipped” hazard ratios for this pre-study ICS subgroup, meaning that the hazard ratio would be oriented as UMEC/VI versus TRELEGY ELLIPTA to account for ICS removal versus ICS continuation, similar to WISDOM and other studies.

If we turn our attention to the right panel for a moment, data from the TRELEGY ELLIPTA versus UMEC/VI comparison in this ICS-naïve subgroup provides information on the effect ICS addition, while keeping all other factors equal. Efficacy data in this subgroup could be considered the “cleanest” data to support a claim of a benefit on all-cause mortality with the addition of fluticasone. In IMPACT, this ICS-naïve subgroup is underpowered to detect a difference in mortality, and the results should be interpreted with caution, but even a trend in this subgroup would help support the idea that the overall analysis result was not driven primarily by effects among subjects who underwent ICS removal. Unfortunately, the data for this ICS-naïve subgroup does not show a trend towards improved all-cause mortality for TRELEGY ELLIPTA. Indeed, at study end the FF/UMEC/VI arm has the numerically highest proportion of deaths. Perhaps equally importantly – in this right panel where ICS removal events could not occur – we see no evidence of an early separation within the first 90 days in the UMEC/VI curve.

There are a few last points to note about the curves on the left. First, I want to again highlight the early take-off of the green UMEC/VI curve within approximately 90 days of ICS removal at randomization; after this initial 90 days, the curve runs a nearly parallel course to the blue and red curves, similar to what was noted in the statistical presentation of IMPACT’s overall ACM result. It’s also worth noting that the red FF/VI arm – which also contains an ICS and could not have ICS removed – follows a similar course to the TRELEGY ELLIPTA arm, while the UMEC/VI arm is remarkably different.

## Slide 15

To examine these subgroup data further, here is a table of all-cause mortality results at different timepoints in the pre-study ICS subgroup. If we look at the subgroup data without respect to ICS-removal events and consider FF/UMEC/VI versus UMEC/VI, the all-cause mortality analysis at 52 weeks appear to suggest a mortality benefit for TRELEGY ELLIPTA compared to UMEC/VI attributable to fluticasone furoate, the ICS component. The same could be said of the data at 90 days. But this interpretation relies on a comparison that is oriented towards TRELEGY ELLIPTA as the intervention, and

UMEC/VI as the control. In the last few slides, we discussed that, for this pre-study ICS subgroup, one could consider ICS-removal to be the intervention and ICS-continuation to be the control. So, to evaluate the effect of ICS removal using this “flipped” orientation for the comparison, we also have to flip the orientation of the hazard ratios. These “flipped” hazard ratios then allow us to characterize the way that ICS removal may have affected all-cause mortality in IMPACT. If you do this, as is presented here, the interpretation changes as well. With this orientation, the data from the UMEC/VI versus FF/UMEC/VI comparison suggest a hazard ratio of 1.64 for all-cause mortality at 52 weeks attributable to ICS removal.

Even more concerning are the all-cause mortality data at 90 days in this pre-study ICS subgroup. These are exploratory analyses where the 90-day time period of evaluation was in part data-driven and therefore may be subject to bias, and there is considerable uncertainty around the estimates due to the small numbers of events. Nevertheless, the results are striking, with a point estimate that would suggest a potentially clinically significant, five-fold increased risk of mortality attributable to ICS removal in this timeframe.

#### **Slide 16**

This potential safety signal for early mortality due to ICS removal in the pre-study ICS subgroup is concerning from a clinical perspective. In an effort to understand this early mortality difference and the perspective on these data during the trial, the Division requested the meeting minutes of the Independent Data Monitoring Committee for IMPACT.

As detailed in your briefing document under Section 4.5.5, these meeting minutes, quoted here verbatim using italicized text, suggest that the early mortality events were noted by the data monitoring committee, and that they were particularly concerned with them. The meeting minutes also suggest that the possibility of inhaled corticosteroid removal was considered in this context. The meeting minutes suggest that a protocol review was undertaken, and that the IDMC found that inhaled steroid use was prohibited 30 days prior to screening and during the study. As we saw in my earlier presentation, however, it was only systemic corticosteroids that were prohibited for 30 days prior to screening. Inhaled corticosteroids were allowed up to randomization, and were also part of the randomized therapies in IMPACT, so ICS removal could – and did – occur.

In addition, we have seen that data from ICS removal trials may suggest that prior exacerbation history could influence the effect of ICS removal in COPD. However, the meeting minutes from that same meeting suggest that complete data on prior exacerbation history were not available to the IDMC at the time of this ICS removal concern. This quote from the meeting minutes (with names replaced) suggests that there were a large number of subjects missing data for prior exacerbations at that time.

#### **Slide 17**

If, on the other hand, we turn our attention back to the data from the ICS-naïve subgroup, using the “standard” orientation of the TRELEGY ELLIPTA versus UMEC/VI, we can see that the confidence intervals are wide on each measure, and we are not able to make definitive statements. However, the data do not suggest a trend towards benefit for TRELEGY ELLIPTA compared to UMEC/VI.

As I mentioned before, however, these data should be interpreted with caution since this exploratory subgroup is not adequately powered to detect a mortality signal.

## Slide 18

As previously discussed, the COPD literature suggests an association between severe COPD exacerbations and mortality, and some authors have suggested a causal role for severe exacerbations as drivers of COPD mortality. To better understand the mortality data, we requested additional analyses of the time-to-first severe exacerbation by pre-study ICS subgroup.

These data are presented here as Kaplan Meier curves for the probability of first severe COPD exacerbation through week 52 for the pre-study ICS and ICS-naïve subgroups. As before, the pre-study ICS subgroup is pictured here in the left panel, while the right panel represents the results in the ICS-naïve subgroup. Again, the TRELEGY ELLIPTA arm is the blue curve, and the UMEC/VI arm is the green curve. The FF/VI curve, in red, is included for completeness. The x-axis is again time, and the y-axis is the probability of first severe COPD exacerbation. In this visualization, the y-axis spans 0 to 20 percent.

If we look at the severe exacerbation results for this pre-study ICS subgroup in the left panel, we can see that the UMEC/VI arm and the TRELEGY ELLIPTA arm again suggest an early division that evolves in approximately the first 90 days. But, just as we did for all-cause mortality, if you take the pre-study therapy into account, subjects in the UMEC/VI arm of this subgroup had an ICS removal intervention at randomization, while subjects in the TRELEGY ELLIPTA arm continued ICS and could be considered the control arm. Under this paradigm, these data suggest that ICS removal could lead to earlier severe exacerbation events among these COPD subjects. If we consider that severe exacerbations could contribute to increased mortality events in COPD, these data could help us to understand the mortality results in IMPACT from a clinical perspective.

## Slide 19

With those subgroup data from IMPACT in mind, we can compare the results to SUMMIT and TORCH to understand whether they provide additional support for this paradigm.

## Slide 20

If we look at SUMMIT using the same schematic that we've used before, we can see that ICS removal actually occurred among all subjects in the pre-study ICS subgroup due to SUMMIT's requirement that ICS-containing medications must be stopped prior to enrollment. This occurred before any data collection could begin. Because no data were collected on these patients prior to enrollment, we have no data on the effect of ICS removal from this timeframe. However, we do have some data on these patients during the 4 to 10 day run-in, which are presented in the Appendices of your briefing document.

But, even though the event of ICS removal occurred earlier, we can still use these subgroup data to better understand the effect of randomized ICS removal on ACM by looking at comparisons that isolate the contribution of ICS. The first of these comparisons is the FF/VI versus VI comparison, where the VI arm will represent an ICS removal intervention while the FF/VI arm will represent an ICS continuation control. The second of these comparisons is the placebo versus FF comparison, where placebo represents an ICS removal intervention and FF represents an ICS continuation control. Just like before, while other interventions may also have taken place in these patients, the factorial design means that other trial interventions besides the ICS itself would affect each of the arms within these two comparisons equally.

## Slide 21

If we turn our attention to TORCH using the same schematic, we can again see that – in the pre-study ICS subgroup – ICS removal again occurred among all the subjects in this subgroup at the beginning of the run-in. And while data from the run-in is not randomized, they provide some supporting information for ICS removal effects that is summarized in the Appendices of your briefing document. Just like SUMMIT, the ICS removal occurred earlier, but we can still use these subgroup data to examine the effect of randomized ICS removal. The first of the comparisons that isolates the effect of ICS is the FP/SAL versus SAL comparison, where the SAL arm represents an ICS removal intervention, and the FP/SAL arm represents the ICS continuation control. The second of these comparisons is the placebo versus FP comparison, where placebo represents an ICS removal intervention and FP represents the ICS continuation control.

## Slide 22

If we look at ICS-removal data from all three trials, we see familiar themes. First, I need to note that since we are once again looking at ICS removal in pre-study ICS subgroups, we have again “flipped” the orientations to comparisons that describe ICS removal as the intervention and ICS continuation as the control, and we have oriented the hazard ratios in the same way. I also need to note that ICS removal data at trial end for SUMMIT and TORCH are not included on this slide, but are in your briefing document. As we saw earlier, IMPACT data from the pre-study ICS subgroup suggests an increased risk for death by Day 90 among those who had an intervention of ICS removal compared to those who continued ICS. When we expand this methodology to SUMMIT and TORCH, we see recapitulation of the concerning trend in the Day 90 data, despite the differences in the designs and patient populations of both SUMMIT and TORCH.

We acknowledge that these are data driven, post hoc, exploratory analyses, and may be subject to bias. We also acknowledge that there may be issues with power, and that there are differences in trial designs, enrolled populations, and even fluticasone products between these trials. Nevertheless, in each trial we observed an increased risk of all-cause mortality after ICS removal events at 90 days in comparisons that isolate the effect of the ICS, and quantifying the timeframe and magnitude of ICS removal effects in COPD is not a new concept. With those considerations in mind, these subgroup data may help clinicians understand the safety concerns of removing ICS in patient populations like those represented by IMPACT, TORCH, and SUMMIT and could still provide clinicians with information that could help them avoid clinical decisions that could put patients at risk.

In contrast, we could summarize the effect of fluticasone addition in these trials as follows: Data from the FF/UMEC/VI versus UMEC/VI comparison of the ICS-naïve subgroup in IMPACT, while underpowered, do not suggest improved mortality attributable to the addition of fluticasone furoate. When we expand this methodology to the ICS comparisons in SUMMIT and TORCH, we again see a familiar theme. Here, we see that the effect estimates for nearly all of these fluticasone addition comparisons hover around 1 or higher. These ICS-naïve subgroup data from SUMMIT and TORCH reinforce the ICS-naïve subgroup data seen in IMPACT. And despite these subgroups including over 1500 ICS-naïve subjects from IMPACT, almost 11,000 ICS-naïve subjects from SUMMIT, and almost 3000 ICS-naïve subjects from TORCH, we see that data from all three trials consistently do not suggest that the addition of fluticasone products improved mortality.

We ask the Committee to consider these subgroup results and their interpretation when evaluating the proposed all-cause mortality claim.

**Slide 23**

So, after going through the study designs, the efficacy results, and the subgroup results for all of these trials, it is time to summarize the uncertainties for your consideration.

**Slide 24**

The overall ACM analysis of IMPACT – including all available vital status follow-up data – comparing FF/UMEC/VI versus UMEC/VI yielded a hazard ratio of 0.72 with a 95% CI of 0.53 to 0.99 and a nominal p-value of 0.042. However, IMPACT was a single trial and not designed as a mortality study, and the evaluation of ACM was not performed under strict Type I error control. We acknowledge that all analyses of primary and secondary endpoints met a nominally significant p-value threshold, that mortality is a clinically important outcome, and that many (but not all) of the exploratory analyses of “other” endpoints had nominal p-values below the 0.05 threshold. Nevertheless, the ACM evaluation was one of a long list of exploratory analyses. It would not be unusual to find nominal p-values below 0.05 just by chance when evaluating multiple exploratory endpoints, and such analyses may also be subject to substantial random high bias. In addition, such a mortality benefit attributable solely to the efficacy of fluticasone furoate was clinically unexpected, given our clinical experience with initiating inhaled corticosteroids.

Given these uncertainties, the statistical persuasiveness of the ACM results for the fluticasone furoate component from IMPACT as a stand-alone trial should be considered by the Advisory Committee.

**Slide 25**

In addition to considering the data from IMPACT on its own, we should consider any supportive evidence for an all-cause mortality benefit for fluticasone provided by SUMMIT and TORCH.

**Slide 26**

Due to the uncertainties in our analysis of IMPACT, the Agency examined these two previous trials to provide additional context. SUMMIT and TORCH were longer trials, specifically designed to evaluate mortality, with a larger number of death events, and thus more power and precision to detect a statistical difference in ACM.

In our estimation, these two trials provide the most reliable independent data to help inform the proposed claim based on IMPACT and are considered critical in light of the uncertainties about the persuasiveness of the results from that single study. Both trials did not show a mortality benefit attributable to fluticasone in their ICS comparisons. We acknowledge that these trials involved different populations and different comparisons to evaluate the ICS effect, and that the fluticasone study drug evaluated in TORCH was fluticasone propionate.

Despite these differences, in light of the evidence from these two trials, we ask the committee to consider whether the evidence presented in IMPACT as a single trial supports the claim that addition of fluticasone furoate – as a component of TRELEGY ELLIPTA – improves mortality in COPD.

As a final point on the ACM results across trials, if we think back to the points made by Suissa and colleagues, their analyses suggested an association between a higher proportion of pre-study ICS subjects at enrollment and the observed risk difference for exacerbations at trial end. If this association were to apply to all-cause mortality as well, then data from IMPACT, SUMMIT, and TORCH that isolate the contribution of the ICS could be consistent with what was postulated by Suissa.

#### **Slide 27**

After looking at the ACM data across trials, we can come back to the timeframe of efficacy observed in IMPACT, and what it might mean.

#### **Slide 28**

Time-to-event visualizations show separation of the UMEC/VI arm within approximately the first 90 days. With exclusion of the first 90 days of data, a signal for durable treatment benefit is no longer observed. While these analyses are exploratory and subject to bias, they suggest that the observed difference at trial completion was driven by the early events, that the efficacy of ICS on ACM occurs within the first 90 days, and that little further benefit is gained after the first 90 days. These observations are unexpected, and potentially are not consistent with previous data regarding the efficacy of fluticasone products and other inhaled corticosteroids in COPD.

From a mechanistic standpoint, the efficacy of ICS in COPD may rely on prevention of severe COPD exacerbations. While trials of ICS on ACM in COPD have proposed that prevention of severe AECOPD events could lead to decreased ACM, previous findings do not suggest that such an effect would occur within a 90-day timeframe.

Conversely, previous literature on ICS-removal may suggest an almost immediate clinical deterioration in COPD that persists over time among vulnerable patients with uncontrolled COPD.

We ask the committee to consider the clinical significance and relevance of this observed efficacy timeframe, and whether the early timeframe of these results could be consistent with a mortality benefit attributable to fluticasone furoate, or an early effect of ICS removal on all-cause mortality among COPD patients who already demonstrated uncontrolled COPD despite their pre-study medications.

#### **Slide 29**

Because of this early separation in the UMEC/VI treatment arm in IMPACT, and the protocol design which could result in abrupt removal of ICS in the pre-study ICS subgroup, we need to consider the data for ICS removal from the pre-study ICS subgroup.

#### **Slide 30**

If we look at the pre-study ICS subgroup of IMPACT, the data suggest that the events in this subgroup drove the observed mortality difference.

In IMPACT, those subjects who underwent ICS-removal as an intervention – meaning those randomized to UMEC/VI – had a higher risk of death at Week 52 compared to the subjects that continued ICS in the TRELEGY ELLIPTA arm.

Furthermore, subjects with pre-study ICS randomized to ICS removal demonstrated an increased risk of death by Day 90 compared to ICS continuation. Taking into account uncertainties inherent in these exploratory analyses, the point estimate from IMPACT may suggest a potentially clinically significant, five-fold increased risk of mortality attributable to ICS removal in this COPD patient population from baseline to Day 90. Similar early mortality trends attributable to ICS removal in the pre-study ICS subgroups of SUMMIT and TORCH may suggest that these ICS removal effects may occur even in the setting of trials that employ a run-in period intended as an ICS washout. In addition, examination of severe exacerbations by pre-study ICS suggested an increased risk of severe exacerbations after ICS removal as well.

In contrast, acknowledging limitations due to power and the exploratory nature of these subgroup analyses, a mortality difference attributable to fluticasone addition was not observed in the ICS-naïve subgroups of IMPACT, SUMMIT, or TORCH.

We ask the committee to consider this pre-study ICS subgroup analysis in the context of the trial design that allowed ICS removal, the striking difference in the behavior of the UMEC/VI arm of the pre-study ICS subgroup, the observed interaction p-value of 0.08 between treatment and pre-study ICS status, and previous literature suggesting that an overall analysis combining data from these subgroups may not be interpretable. We also ask the committee to consider whether these results affect the interpretability of results from the overall analysis of IMPACT .

### **Slide 31**

Finally, in light of all these considerations, we have to consider the generalizability of IMPACT's all cause mortality results to clinical practice.

### **Slide 32**

If we consider how the approval of such a labeling claim would affect clinical decision-making, the overall results of the comparison of FF/UMEC/VI vs. UMEC/VI in IMPACT show a HR 0.72, with a 95% Confidence interval of 0.53 to 0.99 as stated in the proposed labeling claim. In principle, the clinical question in practice is whether the addition of fluticasone furoate improves survival over UMEC/VI. On face value, inclusion of these data in labeling could imply an ACM benefit attributable to the addition of fluticasone furoate as a component of TRELEGY ELLIPTA. Without consideration of pre-study ICS subgroups, these data might imply that initiation of ICS therapy in ICS-naïve patients with a history of exacerbations and uncontrolled COPD will improve survival in those patients.

However, 71% of the subjects in the IMPACT trial already had pre-study ICS therapy, and thus could not have had the ICS drug class added as part of the IMPACT trial. As a result, it is unclear whether all-cause mortality data from this large subgroup in IMPACT is relevant to the clinical question of whether the addition of ICS in ICS-naïve patients improves survival. If the observed difference for all-cause mortality for FF/UMEC/VI vs. UMEC/VI was instead driven by ICS removal in uncontrolled patients, these data might represent a very different clinical question, and the proposed labeling claim could be misleading.

### **Slide 33**

On the other hand, if we look at subjects in whom ICS addition was possible, the 29% of subjects in the IMPACT ICS-naïve subgroup are arguably the most relevant data to inform the clinical question of



whether fluticasone furoate – as part of TRELEGY ELLIPTA – improves all-cause mortality. While we acknowledge that this subgroup is underpowered to detect a difference in ACM, the data from the ICS-naïve subgroup of IMPACT do not provide evidence of benefit in the FF/UMEC/VI vs. UMEC/VI comparison, nor do they show a trend towards improved mortality.

And – while we have to acknowledge trial design differences may limit direct cross-study comparisons – if we look to SUMMIT and TORCH to provide us with supportive information for this ICS-naïve subgroup, we still do not see a trend towards a survival benefit in their fluticasone comparisons, despite longer durations and more events.

We ask the committee to consider how the data from these subgroups influence the generalizability of the data from IMPACT to clinical practice, as well as its ability to support the proposed labeling claim.

#### **Slide 34**

Drugs that can improve mortality in COPD are an unmet need, and they represent an important goal uniting clinicians, patients, Industry, and FDA. The IMPACT trial presents an analysis that suggests a mortality difference when comparing TRELEGY ELLIPTA to UMEC/VI, which would suggest an effect of the fluticasone component. Given the importance of a mortality claim and its potential to change practice patterns, however, the data to support such a claim should be robust to scrutiny; we want to get a mortality claim right. TRELEGY ELLIPTA is already approved, and a decision on this supplemental application will not introduce or remove the drug from the market. When we look closer at IMPACT's mortality data, though, there are a number of uncertainties that raise concern about the ability of these data to support a claim of an all-cause mortality benefit for TRELEGY ELLIPTA.

We ask the committee to consider these uncertainties when applying these all-cause mortality results to the context of clinical practice and evaluating the proposed efficacy claim for TRELEGY ELLIPTA.

#### **Slide 35**

This ends the Division presentation of the application. On behalf of the review team and FDA, I would like to thank you for your service on the committee, and for your time. We look forward to your input during the meeting. Dr. Karimi-Shah will now present the Charge to the Committee.

## **August 31, 2020 Pulmonary-Allergy Drugs Advisory Committee Meeting**

### **Script for FDA Presentation 4: Charge to the Committee**

**FDA Presenter: Banu A. Karimi-Shah, MD, Deputy Director, DPACC, OII, OND, CDER, FDA**

#### **Slide 1**

Hello everyone. My name is Banu Karimi-Shah, and I am the Deputy Director of the Division of Pulmonology, Allergy, and Critical Care. On behalf of the Division and the Agency, I would like to take this opportunity to thank all the Advisory Committee members for their participation in this meeting. We look forward to a fruitful discussion on August 31. Now that you have had the chance to view the presentations by Dr. Busch, Ms. Duke, and the Applicant, I would like to take the next 10-12 minutes to provide a brief reminder of the proposed claim, an overview of the scientific issues, the regulatory framework upon which our decision-making is based, and the questions to be discussed and voted upon. These questions will be presented to you again on the day of the meeting.

#### **Slide 2**

The proposed labeling claim is reviewed on this slide.

The claim states that treatment with TRELEGY ELLIPTA reduced the risk of all-cause mortality compared to UMEC/VI based on on- and off-treatment data.

In addition, the Applicant has included a clause stating that 71% of subjects were on ICS therapy at screening, and partially describing some indicators of severity for this subgroup. The proposed labeling goes on to state that FF/UMEC/VI reduced the risk of all-cause mortality in this subgroup compared to UMEC/VI, and further states that the clinical relevance of these results is unknown. The proposed labeling also notes that the evaluation of all-cause mortality was limited by the small sample size among those who were ICS-naïve.

In this claim, which cites a reduction in all-cause mortality of Trelegy (FF/UMEC/VI) vs. UMEC/VI, the Applicant is asserting that the efficacy is attributable to FF (the ICS component).

#### **Slide 3**

As you have heard in the Agency's presentations, there are a number of issues that raise concern about the ability of the results from the IMPACT trial to support an all-cause mortality (ACM) claim for TRELEGY ELLIPTA.

First, there is the statistical uncertainty that arises from the fact that IMPACT was a single trial, and the analysis showing an all-cause mortality difference was not performed under Type I Error control, so the results could be due to chance.

FDA generally requires supportive data under strict Type I Error control from two randomized controlled trials to support an efficacy claim or new indication; otherwise, in some specific settings, a finding of substantial evidence of effectiveness to support a claim can be made based on one trial and confirmatory evidence, a standard which I will review a little later in this presentation.

As IMPACT was a single trial, the Agency looked to two relevant mortality trials in SUMMIT and TORCH to get additional data to support the effect of fluticasone on ACM.

The longer SUMMIT and TORCH trials were specifically designed to evaluate ACM, had a longer duration, and included roughly 3-fold more death events than IMPACT. However, these studies did not provide evidence of mortality effects for the ICS/LABA product in their primary analyses, nor did they provide corroborating evidence of a mortality benefit for fluticasone products in COPD based on comparisons that isolated the effect of fluticasone propionate in TORCH and fluticasone furoate in SUMMIT.

If we then return to IMPACT and examine the timeframe of the observed mortality difference, we see that the Kaplan-Meier curves separated within the first 90 days. If we take this as an early timeframe of efficacy for TRELEGY ELLIPTA compared to UMEC/VI, it would suggest that fluticasone exerts an effect on mortality within about three months. This would be an unexpected result given our familiarity within the inhaled corticosteroid drug class, the inability of SUMMIT and TORCH to demonstrate efficacy for fluticasone over the course of years, and the proposed mechanism for improvement in mortality through prevention of severe COPD exacerbations, which are rare events. However, this early timeframe would not be inconsistent with a timeframe for an effect of ICS removal, described as “immediate and sustained disease deterioration” by one publication. No such early signal was seen in the pre-study ICS naïve subgroup.

#### **Slide 4**

Indeed, if we take pre-study medications into account, the data from IMPACT suggest that subjects with pre-study ICS therapy who had ICS removed by randomization had a clinically significant increased risk of death by Day 90, recalling Dr. Busch’s slides of the “flipped” hazard ratios. A similar trend was observed when examining severe COPD exacerbations. Data from analogous subgroups in SUMMIT and TORCH are similar, further supporting a potential early risk of death attributable to ICS removal. If, on the other hand, we look at the ICS-naïve subjects, where the TRELEGY ELLIPTA versus UMEC/VI comparison might best mimic the clinical decision of whether to add ICS in a patient with uncontrolled COPD, the limited data – while underpowered – do not suggest a mortality benefit for ICS addition. If we try to get additional data from the ICS-naïve subgroups of SUMMIT and TORCH, we find that the analogous ICS comparisons in those trial subgroups also don’t suggest a benefit of ICS addition.

These uncertainties raised by the subgroup analyses could affect the clinical interpretation of the labeling claim and whether these results are generalizable to patients in clinical practice, in which health care providers are considering the benefit of adding a therapy. Seventy-one (71) percent patients in IMPACT entered the study on pre-study ICS, and could be randomized only to ICS removal or ICS continuation (but not ICS addition). It is uncertain whether the trial design of IMPACT was able to answer the clinically relevant question for this TRELEGY ELLIPTA application: namely, whether the addition of fluticasone furoate to UMEC/VI in a subject without previous ICS therapy will decrease all-cause mortality in COPD.

#### **Slide 5**

To frame the discussion, the next few slides will provide an overview of the governing regulations.

FDA's decision to approve an application, whether it be a new drug approval or a labeling claim as is presented with this submission, depends on the determination that the drug meets the statutory

standards for safety and effectiveness, manufacturing and controls, and labeling. The focus of this meeting will be the safety and effectiveness piece of the application. While we have presented mortality as an “efficacy” endpoint throughout our presentations, this endpoint of course addresses both safety and efficacy.

#### **Slide 6**

The efficacy standard in the regulations describes the need for substantial evidence from adequate and well-controlled investigations supporting the language in labeling.

#### **Slide 7**

The regulations governing determinations of effectiveness are further described in guidance documents from the Agency, which were outlined by Ms. Duke in her presentation, where she showed this slide.

The gold standard is evidence from at least 2 adequate and well-controlled studies.

Otherwise, in some specific settings, a finding of substantial evidence of effectiveness to support a claim can be made based on “one adequate and well-controlled clinical investigation plus confirmatory evidence”. Key factors to allow for such a determination include the “persuasiveness of evidence from a single study” and the “robustness of confirmatory evidence.” The guidance indicates that reliance on a single study should “be limited to situations in which the trial has demonstrated a clinically meaningful and statistically very persuasive effect ...” In that light, as pointed out by Ms. Duke, there is often an expectation of evidence of an effect at a statistical significance level considerably lower than 0.05 when a proposed effectiveness claim relies on results from a single study.

#### **Slide 8**

There are a number of safety reasons that could underlie a refusal to approve an application which are summarized here. As all-cause mortality is both an efficacy and safety consideration, I present this standard here for your review and for completeness.

#### **Slide 9**

I will now move on to the Discussion Points and Voting Question.

Question 1 is a discussion question. We ask the committee to discuss the persuasiveness of the data in the IMPACT trial to support the claim that fluticasone furoate, as a component of TRELEGY ELLIPTA, improves all-cause mortality in COPD. We ask that you include the following elements in your discussion:

- a. The exploratory nature of the ACM analysis, the lack of Type I error control, and the strength of evidence in IMPACT
- b. Whether the ACM results from IMPACT are persuasive in light of the additional ACM data from fluticasone comparisons provided by SUMMIT and TORCH
- c. The observed timeframe of the IMPACT results, i.e., the early separation in survival

## Slide 10

Question 2 is also a discussion question. We ask the committee to discuss the implications of pre-study ICS use and ICS-removal on the interpretation of the ACM data in the IMPACT trial. Include the following elements in your discussion:

- a. The clinical understanding of the contribution of ICS to COPD therapy and the effects of ICS removal in patients with uncontrolled COPD and frequent exacerbations
- b. The implications of randomization to study drugs that do not contain ICS among patients with uncontrolled COPD despite pre-study ICS therapy
- c. The observed timeframe of the IMPACT results, i.e., the early separation in survival, and
- d. The pre-study ICS subgroup data from SUMMIT and TORCH, in light of the differences from IMPACT in study design and patient population

## Slide 11

Question 3 is the last discussion question. Here, we ask the committee to discuss the generalizability of the IMPACT data to relevant clinical practice decisions about fluticasone furoate (FF) as add-on therapy in COPD. Please include the following elements in your discussion:

- a. The clinical relevance and persuasiveness of the ACM results from fluticasone comparisons among the ICS-naïve subgroups of IMPACT, SUMMIT, and TORCH
- b. The clinical relevance of data from the pre-study ICS subgroup to inform decisions regarding the addition of FF
- c. The clinical relevance of the IMPACT trial design and its ability to assess the benefit of adding FF, and
- d. The clinical implications of the proposed labeling claim in light of the submitted data

## Slide 12

Finally, Question 4 is a voting question. On the day of the meeting, we will ask that you vote on the following: Do the data from the IMPACT trial provide substantial evidence of efficacy to support the claim that TRELEGY ELLIPTA improves all-cause mortality in patients with COPD?

- a. If no, what further data are needed?

I will remind you that while the absolute votes are recorded, it is your discussion and thought process that are most important to us as we make our final regulatory decision, so as you prepare for the upcoming meeting, your remarks as to why you voted as you did will be most helpful to us.

## Slide 13

On behalf of the team here at the FDA, I would like to express my sincere gratitude to all the panel members for their preparation for this advisory committee. We realize that reviewing both the slides and the briefing document ahead of this virtual meeting was no small undertaking and are greatly appreciative of your time and effort. We look forward to a productive meeting on August 31st. Thank you very much.